

**FRS: DATA MANAGEMENT PROCEDURES DOCUMENT:**  
**Chapter II “System Description”**  
**May 15<sup>th</sup>, 2000 Work-In-Progress**

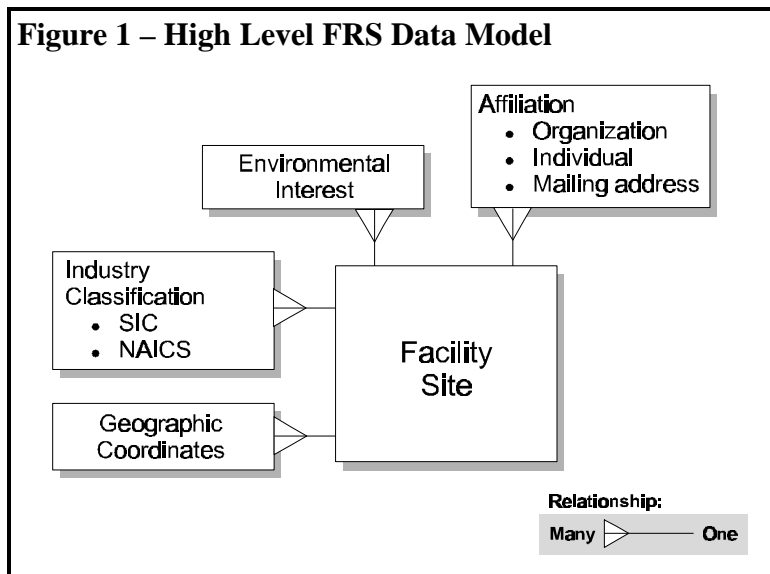
## **II. SYSTEM DESCRIPTION**

The Facility Registry System will serve as the central repository for authoritative facility identification information and will be accessed via a web-based user interface. The FRS will identify each facility with a Facility Registry Identifier (FRI) that will ultimately be used for electronic reporting, central receiving, and integrated reporting. The FRS will provide the capability to identify incomplete data, determine the quality of data, and allow searching and the generation of reports. The FRS will also “link” records across multiple programs or environmental interest areas to a single, distinct facility.

This chapter does three things: (A) characterizes the FRS in terms of its component data elements and the relationships among those data elements; (B) explains the flow of data into the FRS; and (C) describes the functionality of the FRS system.

### **II.A. FRS DATA MODEL**

A complete record in the FRS will consist of 56 distinct data elements. As depicted in Figure 1, data elements fall into five basic categories: facility site, environmental interest, industry classification, affiliation, and geographic coordinates. Figure 1 also outlines the relationships among the five FRS data fields. The five basic FRS data fields and their inter-relationships are described below:



**Facility Site** – A facility site is a distinct real property at a specific place. The facility site data field includes the following elements: facility site name, location address, supplemental location text, locality name, state name, ZIP code, and county name. The FRS data model defines the following relationships between this field and others:

- Each facility site may be involved with one or more *Affiliations*.

- Each facility site may be classified by one or more *Standard Industrial Classification(s)*.
- Each facility site may be classified by one or more *North American Industry Classification(s)*.
- Each facility site may be geographically located by one or more *Geographic Coordinates*.
- Each facility site must be monitored by one or more *Environmental Interest(s)*.

**Affiliation** – This data field characterizes the relationship between the facility site and an organization and/or individual. The affiliation data field consists of the following data elements: affiliation type, organization formal name, organization DUNS number, mailing address, and supplemental address text. The FRS data model defines the following relationships between this field and others:

- Each affiliation occurrence must be established with one and only one *Facility Site* occurrence.
- Each affiliation may involve one and only one *Organization* occurrence; and each organization must be involved with one or more *Affiliations*.
- Each affiliation occurrence may involve one and only one *Individual* occurrence; and each individual must be involved with one or more *Affiliations*.
- Each affiliation occurrence may receive mail at one and only one *Mailing Address* occurrence.

**Geographic Coordinates** – The data elements in this field provide a precise and complete characterization of the physical location of the facility site. The geographic coordinates field includes the following data elements: latitude measure, longitude measure, horizontal accuracy measure, horizontal collection method, horizontal reference datum, reference point, and source map scale. The FRS data model defines the following relationship between this field and others:

- Each geographic coordinates occurrence must geographically locate one and only one *Facility Site* occurrence.

**Industry Classification** – This field contains data elements that characterize the type of business activity occurring at the facility site. The industry classification field includes the following data elements: SIC code, SIC primary indicator, NAICS code, and NAICS primary indicator. The FRS data model defines the following relationship between this field and others:

- Each Standard Industrial Classification occurrence must classify one and only one *Facility Site* occurrence.
- Each North American Industry Classification must classify one and only one *Facility Site*.

**Environmental Interest** – The data elements in this field combine to characterize the need or activity that causes an environmental program to be interested in the facility site. The environmental interest field includes the following data elements: environmental interest type, environmental interest start date, environmental interest start date qualifier, environmental

information system abbreviated name, and environmental information system identification number. The FRS data model defines the following relationship between this field and others:

- Each environmental interest occurrence must apply to one and only one *Facility Site* occurrence.

Table 1, below, provides a definition for each of the 56 FRS data elements.

DATA ELEMENT NAME	DATA ELEMENT DEFINITION
Facility Registry Identifier	The identification number assigned by the EPA Facility Registry System to uniquely identify a facility site.
State Facility Identifier	The unique identification number used by a state to identify a facility site.
Facility Site Name	The public or commercial name of a facility site (i.e., the full name that commonly appears on invoices, signs, or other business documents, or as assigned by the state when the name is ambiguous).
Location Address	The address that describes the physical (geographic) location of the front door or main entrance of a facility site, including urban-style street address or rural address.
Supplemental Location Text	The text that provides additional information about a place, including a building name with its secondary unit and number, an industrial park name, an installation name or descriptive text where no formal address is available.
Locality Name	The name of the city, town, village or other locality, when identifiable, within whose boundaries (the majority of) the facility site is located. This is not always the same as the city used for USPS mail delivery.
County Name	The name of the U.S. county or country equivalent in which the facility site is physically located.
State Name	The name of a principal administrative subdivision of the United States, Canada, or Mexico
Country Name	The name that represents a primary geopolitical unit of the world.

DATA ELEMENT NAME	DATA ELEMENT DEFINITION
Location ZIP Code	The combination of the 5-digit Zone Improvement Plan (ZIP) code and the four-digit extension code (if available) that represents the geographic segment that is a subunit of the ZIP Code, assigned by the U.S. Postal Service to a geographic location.
Tribal Land Name	The name of an American Indian or Alaskan native area where the facility site is located.
Latitude Measure	The measure of the angular distance on a meridian north or south of the equator.
Longitude Measure	The measure of the angular distance on a meridian east or west of the prime meridian.
Horizontal Accuracy Measure	The measure of the accuracy (in meters) of the latitude and longitude coordinates.
Horizontal Collection Method	The code that represents the method used to determine the latitude and longitude coordinates for a point on the earth. The test that describes the method used to determine the latitude and longitude coordinates for a point on the earth.
Horizontal Reference Datum	The code that represents the reference datum used in determining latitude and longitude coordinates. The text that identifies the place for which geographic coordinates were established.
Source Map Scale	The number that represents the proportional distance on the ground for one unit of measure on the map or photo.
Affiliation Type	The name that describes the capacity or function that an organization or individual serves for a facility site.
Organization Formal Name	The legal, formal name of an organization that is affiliated with the facility site.
Individual Full Name	The complete name of a person, including first name, middle name or initial, and surname.
Individual Title Text	The title held by a person in an organization.
Mailing Address	The exact address where a mail piece is intended to be delivered, including urban-style street address, rural route, and PO Box.

<b>DATA ELEMENT NAME</b>	<b>DATA ELEMENT DEFINITION</b>
Supplemental Address Text	The text that provides additional information to facilitate the delivery of a mail piece, including building name, secondary units, and mail stop or local box numbers not serviced by the U.S. Postal Service.
Mailing Address City Name	The name of the city, town, or village where the mail is delivered.
Mailing Address State Name	The name of the state where mail is delivered.
Mailing Address Country Name	The name of the country where the addressee is located.
Mailing Address ZIP	The combination of the 5-digit Zone Improvement Plan (ZIP) code and the four-digit extension code (if available) that represents the geographic segment that is a subunit of the ZIP Code, assigned by the U.S. Postal Service to a geographic location to facilitate mail delivery.
Environmental Interest	The environmental permit or regulatory program that applies to the facility site.
Environmental Interest Start Date	Date the agency became interested in the facility site for a particular environmental interest type.
Environmental Interest Start Date Qualifier	The qualifier that specifies the meaning of the date being used as an approximation for the environmental interest start date.
Environmental Interest End Date Qualifier	The qualifier that specifies the meaning of the date being used as an approximation for the environmental interest end date.
Environmental Information System Abbreviated Name	The abbreviated name that represents the name of an information management system for an environmental program.
Environmental Information System Identification Number	The identification number, such as the permit number, assigned by an information management system that represents a facility site, waste site, operable unit, or other feature tracked by that Environmental Information System.
Standard Industrial Classification (SIC) Code	The code that represents the economic activity of a company (4-digits).
SIC Primary Indicator	The name that indicates whether the associated SIC Code represents the primary activity occurring at the facility site.

DATA ELEMENT NAME	DATA ELEMENT DEFINITION
North American U.S. National Industry Classification System (NAICS) Code	The code that represents a subdivision of an industry that accommodates user needs in the United States (6-digits).
NAICS Primary Indicator	The name that indicates whether the associated NAICS Code represents the primary activity occurring at the facility site.

## II.B. FRS Data Element Business Rules

Business rules are specifications and stipulations that define and preserve the integrity of a data system, such as the FRS. This section articulates and compiles the business rules applicable to each FRS data element.

Some FRS business rules pertain to multiple data elements. The FRS record will be developed from data extracted from Program systems, State Master Files, and/or Central Receiving Certificates. FRS proposes to adopt the following order of priority for data sources: State Master Files, Central Receiving Certificates, TRI, RMP, RCRA TSD, PCS Major, AFS Major, RCRA LQG SF-NPL, PCS Minor, AFS non-Major, RCRA SQG, PADS, and Superfund-CERCLIS. In some cases, FRS data elements may be derived from other sources, such as respected commercial information vendors, such as Dun and Bradstreet. If it is permissible to derive a data element, this is indicated in the listing of business rules. It is also permissible that some FRS data elements contain multiple values; where applicable this is indicated in the listing of business rules.

The following list briefly outlines applicable business rules for each of the 56 FRS data elements.

1. **Facility Site Name:** The following business rules are generic to the stipulation of an FRS site name.
  - This data record can accept multiple values.
  - Facility site names will be standardized by removing special characters, abbreviations, acronyms and anomalies to the extent possible.
  - The FRS name will be a preferred name relying on common English, i.e., all words spelled out and limited acronyms.
  - The standardization process must not change the facility site name used in the source system; only substantive changes in facility name will cause this record to be changed.
  - All names from the source system are maintained behind the FRS name and are searchable.
  - A detailed description of FRS name change rules can be found in Appendix 1.

**2. Location Address:** The following rules apply to this data element.

- P.O. Box is never a valid location address - it is only a mailing address.
- Box by itself is acceptable for rural route location addresses.
- Location addresses will be standardized for validation, verification, and display purposes.
- Display rules will include:
  - Street Suffix Codes will be abbreviated. Ex: St., Rd., Blvd, Ave,
  - Post Directional Codes will be abbreviated. Ex: 1 Maple Ave **SE**
  - Pre-Directional Codes will be spelled out. Ex: 1 **South** Maple Ave
  - Rural route text address will display normal abbreviations. Ex: RR3, Box 2
  - Numbered Streets will follow native format. Ex: 32 23<sup>rd</sup> St NW or 32 Twenty-third St NW
  - Box will always be spelled out.
- Location Addresses can be pulled from sources other than Federal Submissions such as the Web White Pages.

**3. Supplemental Location Text:** The following rules apply to this data element.

- Included only if submitted.
- Can include multiple AKAs.

**4. Locality Name:** The following rules apply to this data element.

- This data record can accept multiple values.
- All must align, i.e., city must be in county, must be in state, must be in Zip.
- FRS will display County that is aligned with location address if facility is located in more than one county.
- This data field must be filled with a valid value from a verifiable source, and will not be derived.

**5. State:** The following rules apply to this data element.

- Must align with locality, zip code, and county.

**Zip Code:** The following rules apply to this data element.

- Must align with locality, state, and county.

**County Name:** The following rules apply to this data element.

- Must align with locality, state, and zip code.

- **Multiple Counties:** The following rules apply to this data element.
  - This data record can accept multiple values.
  - List all counties contained in national systems that are contiguous to primary county name.
  - Change based on direct request from facility.
- 9. **Country Name:** The following rules apply to this data element.
  - Will default to USA unless otherwise marked.
- 10. **Location Description:**
  - The most current value will be the preferred value.
  - This data record can accept multiple values.
- 11. **EPA Region Code:** The following rules apply to this data element.
  - Region must align with State.
  - Can be derived based on Location/Address/State field.
- 12. **Congressional Data:** The following rules apply to this data element.
  - Will be based on lat/long for Location Address. Location address lat/longs can be derived.
- 13. **Legislative District:** The following rules apply to this data element.
  - Will be based on lat/long for Location address.
- 14. **Hydrologic Cataloging Unit (HUC) Code:** The following rules apply to this data element.
  - Derived from Best Value coordinates in Locational Reference Tables.
- 15. **Environmental Interest Status Code:** The following rules apply to this data element.
  - This data element records whether there is an active record in a program system. The rule is if there is an active record in a national system, the Code is “yes.”
  - This data record can accept multiple values.
- 16. **Dun & Bradstreet Number:** The following rules apply to this data element.
  - As reported to a national program system or State Master Record.



- Reconciled by a D&B query.
- 17. Federal Facility Indicator:** The following rules apply to this data element.
- Default is “no” unless flagged in a national program system.
- 18. Tribal Land Indicator:** The following rules apply to this data element.
- Default is “no” unless name exists in Tribal Land name or code field.
- 19. Tribal Land Name [or Code]:** The following rules apply to this data element.
- EPA preferred Source of Tribal Land Names and Codes is AIEO.
- 20. Environmental Information System Abbreviated Name:** The following rules apply to this data element.
- Abbreviations will reflect acronym created by national program system.
  - Acceptable abbreviations will be maintained in permissible value list associated with data element in EDR and maintained in Envirofacts.
  - Business Rules will be enhanced when State Master records are incorporated.
  - This data record can accept multiple values.
- 21. Environmental Information System Identification Number:** The following rules apply to this data element.
- Source of ID numbers are extract files in Envirofacts
  - Only displays “active” system ID numbers. Definition of “active” will depend on definitions in program systems.
  - ID # no longer active in the national systems will also be archived but accessible in the FRS.
  - System ID doesn’t necessarily require contact information.
  - If there is an Information System number, there must be an Information System abbreviated name.
  - Business Rules will be enhanced when State Master Records are incorporated.
- 22. Environmental Interest Type:** The following rules apply to this data element.
- Types will be derived from a list of acceptable values created in collaboration with national program system managers and maintained by OIC.
  - Permissible value list will be accessible in the EDR.
  - This data record can accept multiple values.

- 23. Environmental Interest Start Date:** The following rules apply to this data element.
- Generally the time stamp in the national program system for this ID#.
  - Derived from a list of permissible values based on information in national program systems.
  - This data record can accept multiple values.
- 24. Environmental Interest Start Date Qualifier:** The following rules apply to this data element.
- Derived from a list of permissible values based on information in national program systems.
  - This data record can accept multiple values.
- 25. Individual Full Name - Facility Contact Name:** The following rules apply to this data element.
- This data record can accept multiple values.
  - Public Access point of contact.
  - Can have multiple contact names but only one per national system.
  - Names must be related to an active programmatic record.
  - If program record goes inactive, contact name is archived.
- 26. Individual Title Text - Facility Contact Title Text:**
- 27. Electronic Mail Address:**
- 28. Telephone Number:** The following rules apply to this data element.
- May be derived from sources other than Federal submissions.
  - This data record can accept multiple values.
- 29. Organizational Formal Name:** The following rules apply to this data element.
- Last in, first up if a substantive change occurs.
  - Substantive name changes require manual research.
  - D&B file may be used to validate name (through a query).
  - This data record can accept multiple values.
- 30. Affiliation Type:** The following rules apply to this data element.
- List will be maintained of acceptable values as to type.
  - This data record can accept multiple values.

- 31. Mailing Address:** The following rules apply to this data element.
- All must align street address within city, within ZIP, within State.
  - Mail delivery points are acceptable such as P.O. Box.
  - Same display rules apply to these data elements as apply to Location address data elements.
  - This data record can accept multiple values.
- 32. Supplemental Address Text:**
- 33. Mailing Address City Name:**
- 34. Mailing Address State Name:**
- 35. Mailing Address Country Name:**
- 36. Mailing Address ZIP Code:**
- 37. Parent Company Name:** The following rules apply to this data element.
- Last in, first up if a change occurs.
  - Name change requires manual research.
  - D&B file may be used to validate name (through a query).
- 38. DUNS Company Number:** The following rules apply to this data element.
- To be entered as reported to a national program system or State Master File, and reconciled by a D&B query.
  - This data record can accept multiple values.
- 39. State Business ID:** The following rules apply to this data element.
- To be obtained through the State Report.
- 40. Mailing Address:** The following rules apply to this data element.
- Mailing address elements must align - city/locality, zip, state.
- 41. City Name:**
- 42. State Name:**

- 43. Zip Code:**
- 44. Country:**
- 45. Latitude Measure:** The following rules apply to this data element.
- To be based on “Best Flag Value” in Locational Reference Table from Envirofacts.
  - This data record can accept multiple values.
- 46. Longitude Measure:** See # 45, above.
- 47. Horizontal Accuracy Measure:**
- 48. Reference Point:**
- 49. Source Map Scale:**
- 50. SIC Code:**
- All SIC codes for a facility will be carried.
- 51. SIC Primary Indicator:** The following rules apply to this data element.
- The primary SIC indicator will reflect the primary SICs carried in program systems. There may be more than one Primary SIC indicator because of varying definitions for “primary.”
- 52. NAICS Code:**
- All NAICS codes for a facility will be carried.
- 53. NAICS Primary Indicator:** The following rules apply to this data element.
- Primary NAICS indicator will reflect the primary NAICS carried in program systems. There may be more than one Primary NAICS codes because of varying definitions for “primary.”
- 54. Facility Registry Identifier:** The following rules apply to this data element.
- The Facility Registry Identification does not replace or supercede programmatic identification numbers.
  - A change of zip code or county name will not cause a change in the FRS number.
  - All unique facilities will be uniquely numbered.

- FRS numbers will change only when the linked program interests associated with the location are separated due to whatever circumstance. For example:
  - Ashland Chemical Co has a property with a PCS discharge NPDES permit. It's FRS number is 001. Then Shell Oil buys the entire property and still operates the NPDES pipe, the FRS number does not change. The Facility Site name with change and the data on organization and affiliation but not the FRS number.
  - If Ashland had a NPDES pipe and RCRA storage unit, and Shell bought the plant or area of the business which operated the pipe but not the RCRA storage unit. Then the Ashland FRS record would be re-numbered to 002 and the Shell facility would be assigned a new number 003. The original Ashland FRS record 001 would be archived. Queries on the Shell record would have a time and date stamp with the FRS 003 record and would be related to the associated archived Ashland FRS record 001 as a matter of history. The new Ashland record 002 would also be time and date stamped and would be related to the archived 001 file. By relating the archived records to new records, requesters will always have the opportunity to know that at one time a location had other environmental interests associated with it.

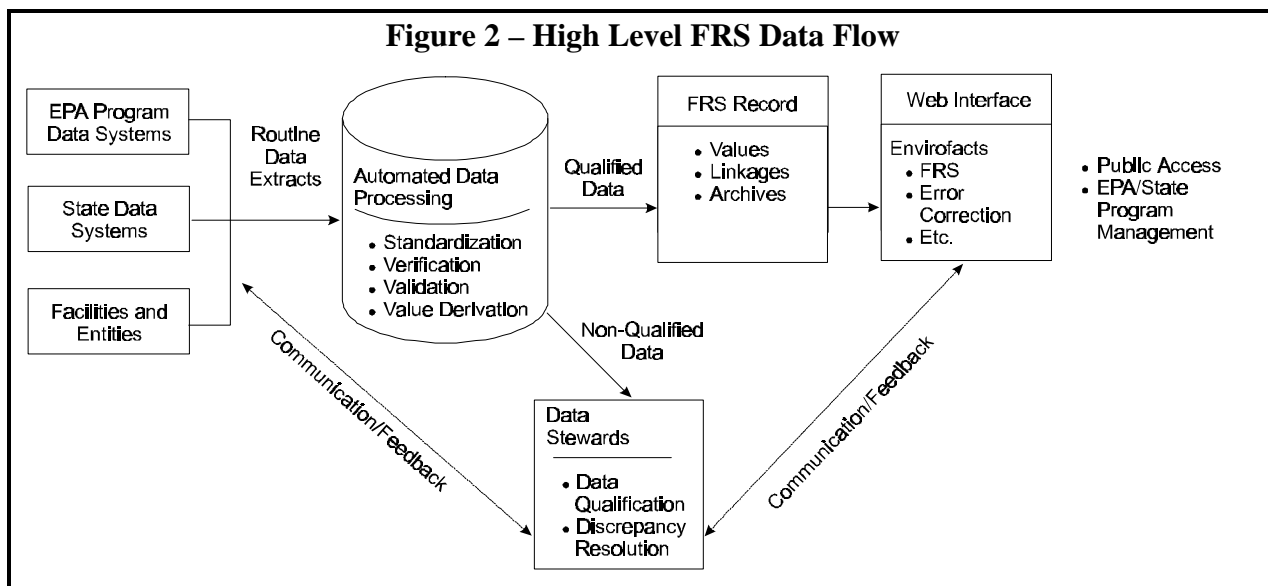
**55. State Facility ID:** The following rules apply to this data element.

- Last in, first up.

**56. FRS Record Creation Date:**

## **II.B. FRS DATA FLOW**

The FRS will obtain facility/site data from a variety of sources including EPA program systems, Central Receiving Registration, State Master Records, and from outside parties. Current records suggest that the Agency has an active interest in approximately 750,000 facilities. Of this number, approximately 50,000 - 60,000 are considered “major” as defined under the various program definitions. Figure 2, below, depicts the basic FRS data flow.



As depicted by Figure 2, the FRS system will extract facility data from State Master Records and various EPA program systems. Facility site name and address data elements are standardized to aid in automated matching processes. Data quality is checked using reference tables. The concept of “linkage” is central to the FRS system. A linkage is an identifier, assigned by an individual media program, that associates a permitted activity or environmental interest with the facility. The “linkage” is the key to obtaining environmental information for the facility from the individual media program system that assigned it.

The basic functionality of the FRS is described below, in Section II.C.

## II.C. SYSTEM FUNCTIONALITY

FRS will contain a single authoritative set of identification data for each facility based on a nationally consistent set of data elements. The FRS system will operate in accordance with stringent data quality requirements and will register only those accurate, current data that completely identify and locate a facility, including accurate linkages to the individual media program systems. A record is said to be “qualified” if it is complete and meets all of the FRS requirements for data quality. A record that does not meet these data quality requirements can become qualified through automated and/or manual derivation of data element values, and/or data steward efforts to augment, validate, verify or derive data values.

The primary functional components of FRS include:

- The ability to create and maintain the authoritative site record and identify the environmental interest for the site.

- The ability to evaluate data quality and report discrepancies to the source of data.
- The ability to identify and verify linkages.
- The ability to generate facility reports.

FRS functions include standardization of key data elements, creation of a unique Facility Registry Identifier (FRI), automated checks of data quality, discrepancy and feedback processing, creation of site records, creation of environmental interest records, facility record search capabilities, and various reporting capabilities. These functions are described below.

**Standardization and Parsing of Data Elements:** The facility site name location address, locality name and name will be standardized and the standardized values will be stored in the facility site table. Location addresses should follow the EPA's address standardization scheme of having street number, street name, and post and/or pre directional references. The location address is parsed into components (i.e., street name, post, and pre-directional) and the components are standardized.

Facility name standardization will be accomplished by using a table-driven approach. The standardized name components will be stored in a table. The standardization process will remove punctuation and special characters from the facility site name. Next, the process will identify components of the facility site name with identical meaning (e.g., TECHNOLOGY, TECHNOLOGIES, TECHNOLOGICAL, TECHNICAL) and replace the component by a common term (e.g., TECH), which is retrieved from the table.

Standardization is also applicable to geographical addresses which may include places with urban-style addresses (e.g., places that can be located by street names and numbers), as well as places that are located by rural routes and highway contract numbers. Address standardization is performed using the table-driven approach. A table stores all the standardized address components. The standardization process removes punctuation and special characters from the location address, it then identifies location address components with identical meaning (e.g., HIGHWAY, HIWAY, HWYS). Identical components are replaced with common terms (e.g., HWY).

The locality and county names are standardized using the table-driven approach where the process starts by converting abbreviated commonly used words such as TWP, CO, or FT to TOWNSHIP, COUNTY, or FORT, respectively. Next, anomalies such as /C/, /T/, /V/, etc., are removed. The standardization process then removes commas and right parenthesis and all the text following either. All other punctuation or special characters and spaces between words are removed.

**Creating a Facility Registry Identifier:** The Facility Registry Identifier (FRI) will be a unique, 12-byte, non-intelligent number based upon a sequence using "linked" numbers between 10,000,000,001 and 19,999,999,999 with a modulus 10 check digit concatenated to its right end (i.e., NNNNNNNNNNNC where NNNNNNNNNNN is the sequence and C is the check digit). The FRI will be stored as a 12-byte field and displayed in the following format: NNNN-NNNN-NNNC.

A new FRI is generated by incrementing an Oracle sequence and a check digit (which is calculated following the guidelines described in the draft FGDC Facility Data Standard) is appended to the generated number to form the complete 12 digit FRI.

**Data Quality Check Procedures:** The FRS conducts an automated data quality review process. Every facility site record in FRS is given a Data Quality Code and a Data Quality Weight. The Data Quality Code is assigned based on the consistency and validity of key facility data elements (facility site name, location address, locality name, county name, state, and postal code). The Data Quality Code is alphabetic and for each core data element there is a corresponding uppercase character. “V” for verified, “N” for site Name, “A” for location address, “C” for locality, “O” for county, “S” for state, and “Z” for location postal code. The lowercase characters in this code are to identify the problems of the core data elements: “i” for invalid/not in conformance with FRS reference tables and “e” for erroneous. The Data Quality Weight is a numeric value based on the quality of the combination of the core data elements’ values, the higher the value the worse the quality of the core data elements.

The standardized facility site name and street name are checked for anomalies (e.g., No Locality Found, No Address, N/A, P.O. Box, Rural Route, Unknown) against an anomalies table. The possible data Quality Code values for facility site name and location address are “eN” (erroneous facility Name) and “eA” (erroneous location address). This means that the value of the site name and/or street name were found in the anomalies table. If values of “eN” or “eA” are assigned to a facility site record, they each generate a Data Quality Weight of 25 points. For example, a Data Quality Code of “eNeA” will generate a weight of 50 points.

The state code is verified first to determine if the system should continue location verification. If the state code is unverified, the process of Data Quality Code assignment is meaningless. In this case, 100 points are added to the Data Quality Weight of the record. If the state code is verified, the standardized locality name, county name, and postal code are evaluated in combination with the state code. The following shows how the Data Quality weight is determined.

- If County name is not verified, 15 points are added to the Data Quality Weight of the record.
- If locality name is not verified, 10 points are added to the Data Quality Weight of the record.
- If postal code is not verified, 5 points are added too the Data Quality Weight of the record.

The locational information is validated using FIPS 55-DC3 and the USPS Zone Improvement Program (ZIP) Code reference tables. The FIPS file was downloaded from the official FIPS web site, which was last updated in 1998. This file provides a two-character state code and five-character numeric place code to uniquely identify each listed entity. An exhaustive list is carried of incorporated places, census designated places, primary county divisions, recognize Indian



reservations and Alaska Native villages; and counties. The USPS ZIP Code file is mainly used to verify postal codes in county boundaries. The locality/postal code boundaries in the USPS ZIP Code file are not as reliable as the county/postal codes. The USPS identifies cities as a post office station in an area, but the USPS city may not necessarily be a true geographic locality name. Therefore, locality name and postal code verification result in lower Data Quality Weights.

**Discrepancy and Feedback Processing:** The facility site records that do not pass the data quality checks will be stored in the feedback processing area of the FRS. The Data Quality Code assigned to each record points to the possible problem data in the core facility data elements. FRS provides capabilities to identify and report on facilities with data quality issues. These reports can be run after each facility update is received by FRS. Copies of these reports will be sent to FRS Data Stewards and/or the source of facility data for correction and re-submission. FRS will track the date when the report was generated and sent to the source for correction and confirmation.

In the feedback processing area, the facility records are uniquely identified by the FRI, source and the last received date. FRS will maintain and populate corresponding tracking dates based on the data quality.

**Creation of Authoritative Site Record:** The initial population of the FRS will be extracted from the TRIS and RMP systems. Initially, the RMP facilities site and the environmental interest records will be loaded to the corresponding resolution tables and processed as follows:

- Facility site records with core data elements missing, or with mailing addresses (i.e., P. O. Box) reported for the physical location of facility will be flagged as incomplete and will not be loaded into FRS.
- Duplicate facilities reported by the same source will be rejected. A facility record is considered duplicate if there is more than one record with the same source program system identification number.
- All incomplete site records and associated environmental interests will be stored in the FRS incomplete facility site and environmental interest tables. Reports reflecting facility information from these tables can be generated and forwarded to the appropriate Data Stewards and/or data sources. Incomplete site records will be tracked using updates received from the source of facility data.
- The site records with complete data will be standardized. Data quality validation rules will be applied to these records and the data that does not pass validation checks will be flagged, and will be recorded in a report to be forwarded to the respective data sources.
- An authoritative site record will be created for every complete RMP site record in the resolution table.

- The data in the resolution tables will be deleted and reloaded with the TRIS facility information. All pre-processing steps mentioned above will apply to TRIS facility site records.
- Reported linkages associated with the TRIS facility site record from resolution table will be compared with the RMP linkages in the environmental interest table. If both programs report the same linkages, the site record will be updated with TRIS facility information and the source will be changed to “TRIS”. Additional TRIS reported linkages will be recorded in the FRS environmental interest table. Source of common linkages will be changed to TRIS.
- If no linkages are matched, data elements of TRIS site records will be compared with the data elements of RMP site records. The algorithm for comparing the data elements and scoring is described below.
  - XVI. State matches.
  - XVII. If location address matches, score = 50.
  - XVIII. If locality name matches, score = 15.
  - XIX. If county name matches, score = 5.
  - XX. If postal code matches, score = 5.
  - XXI. If name matches, score = 25.
- If the TRIS site record in the resolution table satisfactorily matches with one and only one RMP site record in FRS (i.e., location address matches, state matches, and facility name matches), then the FRS site record will be updated with the TRIS information and the source of facility data will be changed to TRIS. The environmental interest records reported by TRIS will be recorded in the FRS environmental interest table under the same facility site record.
- If the TRIS site record in the resolution table reasonably matches with one and only one RMP site record in FRS (i.e., state matches, parts of the location address or complete location address matches, and postal code matches), then the TRIS record will be flagged for further analysis before the creation of the FRS site record.
- If the TRIS site record does not match an RMP site record (name does not match and location address does not match), an FRS site record will be created using the TRIS site record as the source record and a new FRI will be assigned to it. The environmental interest records will be recorded in the FRS environmental interest table.
- FRS site records that pass the data quality validation will be flagged as verified (verified flag= “Y”). Site records that do not pass the data quality validation will be flagged as pending (verified flag = “P”). The registered flag will be set to “Y”, for “Yes”. The facility

interest status code will be set to “A” for “Active” and the facility site type name will be set to “Stationary”.

These pre-processing steps are needed to prevent the creation of duplicate facilities and to ensure high quality data. Due to the variation of facility information from the two different sources, a small percentage of duplicate sites may still exist.

**Creation of Environmental Interest Records:** For every FRS site record, the environmental interest records reported with the site record are loaded into the environmental interest table. The environmental interest data for TRIS and RMP is obtained from the TRIS and RMP records, respectively. Environmental interest data for the other program systems is derived from Envirofacts by matching on the program system identification number. The process for setting the verified flag is as follows:

- The information system identification number is matched with the Program System ID in FLA. If it exists in FLA and if it is linked with the same source, the verified flag of the FRS record will be set to “Y” for “Yes”.
- If the information system identification number is not matched with the program system identification number in FLA, the verified flag in FRS will be set to “N” for “No”. FRS will report all unverified linkages to the facility data source for confirmation and correction.

**Facility Search Components:** The facility search interface will allow users to search the FRS database for facility information. Facility searches can be conducted using multiple search criteria such as geographic information and facility name. Facility search results will display the facility registry identifier, facility name, and locational information. The environmental interest abbreviated name and environmental interest identification number will be displayed on the detailed page.

- Single criterion searches will be limited to:
  - R Facility Registry Identifier.
  - R Facility Name.
- Multiple criteria searches will include:
  - R State in combination with locality, county or postal code.
  - R Postal code in combination with facility name.
  - R Street address in combination with state and locality or postal code.
  - R Facility name in combination with state and locality or county.
  - R Facility name in combination with location address.
- FRS will allow the users to select active versus inactive facilities.

- Facility name searches will provide four matching options: Exact Name, Starting Characters of Name, Sounds Like Name, and Containing.

Facility site data retrieval will provide a list of all facility records where the search criteria are satisfied. The list will be organized alphabetically by facility name. Data displayed on the list will include facility name, FRI, location address, locality, state, postal code, and last update date. From this list, the user will be able to select a facility record to view the detailed facility data.

**Reporting Capability:** FRS will provide an interface to enable users to generate, view, and download facility reports. In addition, reports will also be generated by the various application processes to inform the data various sources about the data problems and linkage discrepancies. The Facility Site Information and Facility Discrepancy reports are part of the web interface. The Incomplete Facility report will be created by FRS background processing. More reports may be developed for future releases.

- *Facility Site Information Report* – This report will display the facility site information for one or more sites; the output will be based on the selected criteria. The following options will be available on this report:
  - A report type of the Facility Summary Report or a Facility Detailed Report can be selected. The summary report displays the FRI, facility site name, location address, locality name, county name, state name, postal code, and the verified flag. The detailed report displays the above elements plus the associated environmental interest information.
  - The report allows search selection criteria on FRI, facility site name, location address, locality name, county name, state name, postal code, and the verified flag.
  - The report allows the selection of the order in which the core site information and other associated information are displayed on the report output. This function also allows the selection of which site information to display. If no site information is selected, a default report with all the site information will be generated.
  - Three ascending sort options can be selected and the capability to view the information on-line or download the results in fixed or comma delimited format is available.
- *Facility Discrepancy Report* – This report will list all the facilities which do not pass the data quality validations. The FRI, facility site name, location address, locality name, county name, state name, postal code, and description of the data problem will be displayed in the results. The report can be reviewed online or downloaded in a fixed format. If the report is viewed online, the site information that did not pass the validation will be highlighted in color. If the report is downloaded, the site information that did not pass the validation will be surrounded by brackets.

- *Incomplete Facilities Report* – This report will list all the facilities which are identified as incomplete (the core data elements are missing or a P.O. Box is in the location address). The report is produced by FRS background processing; it is not an application interface report. The FRI (if provided), facility site name, location address, locality name, county name, state name, postal code, associated environmental interest, and source of data will be displayed in the output set. This report will be forwarded to the reporting source so that complete data can be reported in the future.

**History of Facility Site and Related Information:** The FRS application will maintain historical information identifying the changes to the facility data. When changes are made to the facility site data, a historical record will be created in the Facility Site history table with an exact copy of the previous record. Updates to the environmental interest information will not be captured in a history table; the existing information will be replaced by most recent update from the source of data.